

# UNE ÉVALUATION DE LA QUALITÉ DES DONNÉES À L'ENQUÊTE SUR LA SANTÉ DANS LES COLLECTIVITÉS CANADIENNES

*Yves BÉLAND, Johane DUFOUR et Lyne BÉRUBÉ*

*Statistique Canada – Division des méthodes des enquêtes auprès des ménages*

## **Introduction**

En 1998, le Conseil consultatif sur l'infrastructure de la santé, l'Institut canadien d'information sur la santé (ICIS) et Statistique Canada ont réuni plus de 500 personnes pour mettre en place une stratégie permettant d'obtenir une meilleure information sur le système de santé au Canada. Les consultations auprès d'experts ont mené à la création de *l'Initiative du Carnet de route de l'information sur la santé* (ICIS, 1999) qui présente une vision canadienne commune pour moderniser et normaliser le système d'information sur la santé dans les diverses régions du pays.

Dans le cadre de cette initiative fédérale, Statistique Canada a mené, au printemps 1999, des consultations à travers le pays auprès de plus de 225 intervenants clés du domaine de la santé afin de mieux cerner les besoins en matière de données. Ces consultations ont permis d'établir certaines exigences qui ont dicté le développement méthodologique de l'Enquête sur la santé dans les collectivités canadiennes (ESCC).

L'ESCC, dont les activités de collecte ont débuté en septembre 2000, comprend deux enquêtes transversales distinctes formant un cycle bisannuel répétitif (Béland, Bailie, Catlin et Singh, 2000). La première enquête du cycle, appelée composante régionale, est conçue pour recueillir des données auprès d'un échantillon de plus de 130 000 répondants pour produire des statistiques pour 136 régions sociosanitaires. La deuxième enquête (composante provinciale) est conçue pour fournir des estimations à l'échelle provinciale sur un thème particulier à partir d'un échantillon de 30 000 répondants ; le thème qui a été identifié pour le premier cycle est la santé mentale et le bien-être.

En mai 2002, Statistique Canada diffusait des données détaillées sur la santé pour 131 535 Canadiens âgés de 12 ans et plus, répartis dans 136 régions sociosanitaires (RSS).

Cet article dresse une évaluation de la qualité des données de cette enquête majeure. La section 2 traite du plan d'échantillonnage et des activités de collecte et de traitement de cette enquête pour bien situer le contexte. La section 3 procure des renseignements quantitatifs sur des sources précises des erreurs pouvant s'être glissées dans les données de l'ESCC. La qualité de la couverture des bases de sondage, le suivi des opérations de collecte, les erreurs dues à la non-réponse, l'incidence du mode de collecte quant aux erreurs de réponse (interviews personnelles versus téléphoniques) et la forme de

*Une évaluation de la qualité des données à l'enquête sur la santé dans les collectivités canadiennes* 3

déclaration (auto-déclaration versus par procuration), la performance du modèle d'imputation et les erreurs d'échantillonnage sont abordés.

## **1. Description de l'enquête**

L'objectif premier du premier cycle de l'ESCC vise d'abord et avant tout à fournir des estimations transversales fiables à l'échelle de la région sociosanitaire en ce qui a trait aux déterminants de la santé, à l'état de la santé et à l'utilisation du système de santé.

### **1.1 Population cible**

La composante régionale de l'ESCC ne visait que les personnes vivant dans des logements privés qui sont âgées de 12 ans et plus. Étaient exclus du champ de l'enquête les personnes qui vivent sur les réserves indiennes et sur les terres de la Couronne, les résidents d'un établissement institutionnel, les membres à temps plein des Forces armées canadiennes et les résidents de certaines régions éloignées. L'ESCC couvrait environ 98 % des 12 ans et plus de la population canadienne.

### **1.2 Régions sociosanitaires**

Pour des raisons administratives, chaque province est divisée en régions sociosanitaires (RSS) et chaque territoire nordique est considéré comme une RSS unique (tableau 1). En collaboration avec les autorités provinciales, Statistique Canada a apporté des modifications mineures aux limites de certaines RSS de sorte qu'elles coïncident avec la géographie du Recensement de la population de 1996. La composante régionale de l'ESCC est conçue pour recueillir des données sur 133 RSS réparties dans les 10 provinces, en plus d'une RSS par territoire, donc au total 136 RSS.

### **1.3 Taille et répartition de l'échantillon**

Afin de produire des estimations précises pour les 136 RSS et compte tenu du budget accordé à l'ESCC, il a été établi qu'un échantillon de 133 300 personnes était nécessaire pour atteindre les objectifs. Bien que la production d'estimations fiables à l'échelle de la RSS soit l'un des principaux objectifs de l'enquête, la qualité des estimations provinciales produites pour certaines variables importantes était également considérée comme étant essentielle. Par conséquent, la stratégie de répartition de l'échantillon, qui compte trois étapes, a accordé à peu près le même poids aux RSS et aux provinces. Aux deux premières étapes, l'échantillon a été réparti entre les provinces et les territoires proportionnellement à leur population respective et au nombre de RSS qu'ils contiennent (tableau 1). À la troisième étape, l'échantillon a été réparti entre les RSS proportionnellement à la racine carrée de la population estimative de chacune d'elles. Cette méthode en trois étapes a permis d'attribuer une taille d'échantillon suffisante à chaque RSS en perturbant le moins possible la répartition à l'échelle provinciale. La taille d'échantillon a été augmentée avant la collecte des données afin de tenir compte des unités hors du champ de l'enquête, ainsi que du taux prévu de non-réponse. (Pour la liste complète des RSS et des tailles d'échantillon, consulter : [http://www.statcan.ca/enquetes\\_sante](http://www.statcan.ca/enquetes_sante).)

### **1.4 Bases de sondage et stratégies d'échantillonnage des ménages**

L'échantillon de ménages de l'ESCC a été sélectionné à partir de trois bases de sondage. La majorité de l'échantillon (environ 83 %) provient d'une base de sondage aréolaire développée à Statistique Canada principalement pour l'Enquête sur la population active (EPA). Pour certaines RSS et pour certains mois de collecte, on a utilisé une base de sondage à composition aléatoire (CA) ou une liste de

numéros de téléphone. Environ 7 % de l'échantillon de ménages provenait de la base de sondage à CA et presque 10 %, de la base liste de numéros de téléphone.

La sélection des logements dans la base aréolaire a été effectuée selon un plan d'échantillonnage en grappes stratifié à plusieurs degrés (Gambino, Singh, Dufour, Kennedy et Lindeyer, 1998). Lors de la première étape de l'élaboration du plan de sondage, on a dressé une liste des logements à l'intérieur des grappes sélectionnées. À la deuxième étape, on a sélectionné un échantillon de logements dans chacune des grappes d'après cette liste. Les ménages vivant dans les logements sélectionnés ont ainsi formé l'échantillon de ménages. Des personnes ont été sélectionnées aléatoirement dans les ménages ainsi échantillonnés en vue de participer à une interview personnelle. (Il est à noter que certaines interviews ont eu lieu au téléphone puisque dans plusieurs cas la personne sélectionnée n'était pas à la maison au moment de la visite de l'intervieweur.)

**Tableau 1 : Nombre de régions sociosanitaires et tailles d'échantillon selon la province/territoire**

Province	Nombre de RSS	Taille totale de l'échantillon visée (répondants)
Terre-Neuve	6	4 010
Île-du-Prince-Édouard	2	2 000
Nouvelle-Écosse	6	5 040
Nouveau-Brunswick	7	5 150
Québec	16	24 280
Ontario	37	42 260
Manitoba	11	8 000
Saskatchewan	11	7 720
Alberta	17	14 200
Colombie-Britannique	20	18 090
Yukon	1	850
Territoires du Nord-Ouest	1	900
Nunavut	1	800
<b>Canada</b>	<b>136</b>	<b>133 300</b>

L'échantillonnage des ménages à partir de la base de sondage à CA a été réalisé selon la méthode d'élimination des banques non valides (EBNV) adoptée par Statistique Canada pour l'Enquête sociale générale (Norris et Paton, 1991). Pour les besoins de l'échantillonnage, on a considéré qu'une banque de numéros de téléphone (définie par un indicatif régional de trois chiffres et les cinq premiers des sept chiffres d'un numéro de téléphone) était « valide » si elle comprenait au moins un numéro de téléphone résidentiel. Les banques valides ont été regroupées pour former des strates CA englobant, de façon aussi proche que possible, les régions sociosanitaires. Dans chaque strate CA, une banque a été sélectionnée au hasard pour ensuite générer un nombre aléatoire entre 00 et 99 pour créer un numéro de téléphone complet à 10 chiffres. Les personnes sélectionnées dans les ménages provenant de cette base ont été interviewées au téléphone.

Pour la majorité des RSS, l'échantillon de ménages provenant de la base liste a été obtenu en effectuant un échantillonnage aléatoire simple parmi la liste de tous les numéros disponibles dans la RSS. Cette liste avait préalablement été créée en couplant l'annuaire téléphonique du Canada, un cédérom disponible sur le marché, à des fichiers administratifs internes de conversion afin d'obtenir les codes postaux que l'on a fait correspondre aux RSS. À l'instar des unités tirées à partir de la base à CA, les personnes sélectionnées dans les ménages provenant de la base liste ont été interviewées au téléphone.

## 1.5 Sélection des répondants

Dans 82 % des ménages échantillonnés à partir de la base aréolaire, on a sélectionné au hasard une personne de 12 ans et plus dans chaque ménage, et pour les 18 % restants, on a sélectionné au hasard deux personnes de 12 ans et plus. Cette sélection a été faite en fonction de la composition du ménage afin d'augmenter la représentation des deux groupes d'âge présentant un intérêt particulier, à savoir les jeunes (de 12 à 19 ans) et les personnes âgées (65 ans et plus). Dans les ménages échantillonnés à partir des deux bases téléphoniques, on a sélectionné aléatoirement une personne de 12 ans et plus parmi l'ensemble des membres du ménage.

## 1.6 Collecte des données, méthodes d'interview, suivi de la non-réponse

La collecte des données de l'ESCC s'est étalée sur 14 mois, soit de septembre 2000 à novembre 2001, afin d'équilibrer la charge de travail des intervieweurs et de réduire au minimum les effets saisonniers éventuels sur certaines caractéristiques. Tous les ménages sélectionnés ont reçu une lettre de présentation et un dépliant avant même la visite ou l'appel d'un intervieweur expliquant l'importance du sondage et décrivant quelques exemples d'utilisation des données recueillies. Pour l'ESCC, toutes les interviews, qu'elles soient en personne ou par téléphone, ont été réalisées selon la méthode d'interview assistée par ordinateur (IAO) et tous les intervieweurs ont été formés en conséquence. En plus, afin d'assurer un bon déroulement des activités de collecte, un programme d'assurance de la qualité comprenant une structure de surveillance et de contrôle des intervieweurs a été en application durant la collecte.

À cause de la nature personnelle et délicate de bon nombre de questions de l'ESCC, les interviews par procuration n'étaient autorisées que pour les personnes sélectionnées qui i) étaient absentes pour toute la durée de l'enquête ; ii) avaient une incapacité physique et/ou mentale ou iii) ne pouvaient communiquer avec l'intervieweur à cause d'une barrière linguistique.

Lors de la collecte, les intervieweurs avaient reçu comme instructions de faire toutes les tentatives raisonnables pour réaliser les interviews nécessaires à l'ESCC. Lorsque la visite de l'intervieweur tombait à un mauvais moment, il prenait un rendez-vous à un moment plus convenable. S'il n'y avait personne à la maison, il effectuait de nombreuses visites de rappel. Aux personnes qui refusaient d'entrée de jeu de participer à l'ESCC, Statistique Canada envoyait une nouvelle lettre soulignant l'importance de l'enquête et de la collaboration du ménage. Suivait une seconde visite, d'un intervieweur principal, d'un surveillant de projet ou d'un autre intervieweur qui tentait de convaincre les répondants de participer à l'enquête. Au cours des derniers mois de la collecte des données, on a revu les cas de non-réponse et contacté les répondants choisis qui avaient précédemment refusé pour les encourager à participer à l'enquête. Pour pallier le problème de langue susceptible de nuire aux interviews, tous les bureaux régionaux de Statistique Canada ont embauché des intervieweurs qui parlaient plusieurs langues. Lorsqu'il le fallait, des cas ont été transférés à un intervieweur apte à remplir le questionnaire dans la langue voulue. De plus, les questions de l'enquête ont été traduites dans les langues suivantes : le mandarin, le panjabi et l'inuktitut.

## 1.7 Conception du questionnaire

À la phase de développement du questionnaire, les membres de l'équipe ont consulté un grand nombre d'utilisateurs de données sur la santé de l'ensemble du Canada. Ces consultations ont permis d'évaluer les besoins en données et d'élaborer le contenu du questionnaire de l'ESCC. L'un des objectifs importants de l'ESCC était de recueillir des données sur des sujets ayant une importance particulière pour les RSS. Pour atteindre cet objectif, le questionnaire a été divisé en deux parties, à savoir une

partie à contenu commun d'une durée de 35 minutes et une partie à contenu optionnel d'une durée de 10 minutes contenant des questions offertes aux RSS pour répondre à leurs besoins particuliers. Cette stratégie a produit 27 versions distinctes du questionnaire. Le questionnaire complet de l'ESCC est disponible sur le site Web de Statistique Canada à : [http://www.statcan.ca/enquetes\\_sante](http://www.statcan.ca/enquetes_sante).

## **1.8 Traitement des données, imputation en cas de réponse par procuration**

Une application IAO permet d'effectuer bon nombre de vérifications des données au moment de la collecte. Dans le cadre de l'ESCC, l'application IAO ne permettait pas la saisie de valeurs tombant hors des fourchettes établies et contrôlait les erreurs d'enchaînements des questions. L'application s'assurait aussi que seules les questions qui s'appliquaient à une personne particulière lui soient posées. Dans d'autres situations, l'application affichait des messages d'avertissement. Seules quelques incohérences ont été détectées et résolues lors des vérifications réalisées par les employés du bureau central.

À cause de leur nature personnelle ou délicate, bon nombre de questions de l'ESCC se prêtaient uniquement à l'auto-déclaration et n'étaient pas posées lorsque les renseignements étaient recueillis auprès d'une personne qui répondait par procuration. Une proportion plus forte que prévue d'interviews a été réalisée par procuration (6,3 % au lieu d'une proportion anticipée de 2-3 %), si bien que des renseignements importants manquaient pour les personnes visées (cette proportion pouvait atteindre 23 % dans certaines RSS). Par conséquent, les valeurs manquantes ont été imputées par la méthode du « plus proche voisin » à l'étape du traitement des données (St-Pierre et Béland, 2002). L'imputation n'a été utilisée que pour compléter les renseignements recueillis sur les personnes pour lesquelles l'interview avait été réalisée par procuration. Elle n'a pas été utilisée en cas de non-réponse totale ou partielle lors des interviews réalisées auprès de la personne sélectionnée proprement dite.

## **1.9 Pondération et variance**

Un poids d'échantillonnage a été attribué à chaque personne interviewée afin de représenter sa contribution au total de la population. Comme l'échantillon de l'ESCC provenait de bases de sondage chevauchantes, selon des plans d'échantillonnage distincts, deux méthodes de pondération comportant divers ajustements ont été suivies en parallèle, puis intégrées selon une technique appliquée aux bases de sondage doubles (bases aréolaire et téléphoniques). Puis, on a calé les poids intégrés sur les totaux projetés de population par stratification unidimensionnelle a posteriori selon 10 strates âge-sexe dans chaque RSS (Brisebois et Thivierge, 2001).

En raison de la complexité du plan d'échantillonnage et de la méthode d'estimation de l'ESCC, il est évidemment impossible de se servir de formules de variance de base pour estimer correctement les erreurs d'échantillonnage à l'aide des données de l'enquête. La méthode de rééchantillonnage *Bootstrap* a été utilisée afin de permettre aux utilisateurs d'obtenir des mesures de précision. On a commencé par tirer aléatoirement 500 sous-échantillons *Bootstrap* à partir de l'échantillon principal pour ensuite pondérer chacun d'entre eux avec toutes les corrections nécessaires (ajustements de la non-réponse, intégration, calage sur marges, etc.). L'ensemble des 500 poids *Bootstrap* pour chaque répondant de l'enquête a été fourni, avec instructions et exemples d'utilisation, à tous les utilisateurs des données de l'ESCC pour qu'ils puissent calculer leurs mesures de précision.

## **1.10 Diffusion des données**

En mai 2002, Statistique Canada diffusait les premières données de l'ESCC. Cette diffusion comprenait différents produits, notamment des indicateurs de santé régionaux et certains rapports analytiques disponibles sur le site Web de Statistique Canada ; des fichiers de données spécifiques (partagés et appariés) pour nos partenaires (ministères provinciaux de la Santé et Santé Canada) respectant les restrictions sur le partage de l'information ou appariant les données d'enquête à d'autres sources d'information sur la santé, un fichier maître à l'intention des utilisateurs des Centres de données sur la recherche de Statistique Canada, un fichier microdonnées à grande diffusion à l'intention du public ainsi qu'un fichier de données fictives nécessaire aux utilisateurs du programme de télé-accès.

## 2. Évaluation de la qualité des données

Dans le but d'assurer et de contrôler la qualité des données de l'ESCC, un système de gestion de la qualité a été mis sur pied. Durant la collecte, sur une base mensuelle ou trimestrielle, toute une variété d'indicateurs de qualité a été produite et analysée (taux de non-réponse, taux de vacance, taux de réponse par procuration, tailles observées, etc.). D'autres indicateurs ont été minutieusement analysés sur une base moins régulière ou uniquement à la fin de la collecte des données (taux de glissement, couverture téléphonique, taux d'imputation, effets de plan de sondage, etc.).

### 2.1 Qualité de la couverture

Les erreurs de couverture se produisent lorsque les unités d'échantillonnage de la base de sondage ne représentent pas convenablement la population cible au moment de l'enquête. Il se peut que des unités aient été omises de la base de sondage (sous-dénombrement), que des unités ne se trouvant pas dans la population cible y aient été incluses (surdénombrement) ou que des unités s'y trouvent plus d'une fois (répétitions). À Statistique Canada, l'indicateur de qualité utilisé pour mesurer l'erreur de couverture s'appelle le *taux de glissement*. Par définition, ce taux représente le pourcentage d'écart entre les estimations démographiques de l'ESCC (sans données externes) et les plus récentes estimations démographiques basées sur le recensement. On tient compte de l'erreur de couverture par la pondération des estimations de l'ESCC selon différents groupes d'âge-sexe à l'échelle des régions de santé (Brisebois et Thivierge, 2001).

Les tableaux 2.1.1 et 2.1.2 présentent respectivement les taux de glissement pour les provinces/territoires et par groupe d'âge-sexe à l'échelle nationale. Pour le Canada, le taux de glissement s'établit à 10,5 %. Ce taux s'apparente au taux observé par d'autres enquêtes de l'Agence, et plus principalement à celui de l'EPA d'où provient la majorité de l'échantillon (83 %). Les taux provinciaux varient entre 6,7 % et 16,1 % ; encore une fois ces taux ne sont pas inhabituels. Les trois territoires canadiens affichent des taux supérieurs au taux national. Les taux de glissement par groupe d'âge-sexe (tableau 3.1.2) fluctuent entre 2,2 % (observé chez les hommes de 65 ans et plus) et 21,9 % (enregistré chez les hommes de 20-29 ans). Ces deux résultats s'inscrivent au nombre des résultats bien connus pour les enquêtes de Statistique Canada : les hommes canadiens de 20 à 29 ans constituent une population mobile difficile à dénombrer ; les Canadiens plus âgés sont plus sédentaires et en conséquence plus facile à énumérer.

**Tableau 2.1.1 : Taux de glissement par province**

Province	Taux de glissement
Terre-Neuve	16,1
Île-du-Prince-Édouard	11,3
Nouvelle-Écosse	10,6
Nouveau-Brunswick	13,3

Québec	8,9
Ontario	10,3
Manitoba	6,7
Saskatchewan	9,4
Alberta	7,8
Colombie-Britannique	15,5
Yukon	24,2
T.N.-O.	20,7
Nunavut	20,8
<b>Canada</b>	<b>10,5</b>

**Tableau 2.1.2 : Taux de glissement national par groupe d'âge-sexe**

Âge	Homme	Femme
12-19	9,9	5,8
20-29	21,9	13,0
30-44	17,1	8,6
45-64	10,7	6,0
65 +	2,2	5,8

L'ESCC combine l'utilisation de trois bases de sondage et de deux modes de collectes, soit des interviews en personne pour l'échantillon sélectionné de la base aréolaire (83 %) et des interviews téléphoniques pour les unités échantillonnées de la base à CA (7 %) et de la base liste de numéros de téléphone (10 %). La base à CA et la base liste de numéros de téléphone ne permettent pas de couvrir l'ensemble de la population cible. En effet, le tableau 2.1.3 présente des statistiques (moyennes, minima, maxima) sur le taux de couverture téléphonique par région sociosanitaire à l'échelle provinciale et nationale. En moyenne, 98 % de la population canadienne (excluant les territoires) est couverte par l'utilisation du téléphone. La base aréolaire permet l'utilisation de ces deux bases puisqu'elle pallie le 2 % manquant en couvrant tous les logements privés et en faisant appel à des interviews en personne.

Dans le cas de la base liste de numéros de téléphone, une partie plus importante de la population n'est pas couverte ; elle couvre en moyenne près de 80 % de la population cible (tableau 2.1.4). En effet, tous les numéros confidentiels ou non publiés, les ménages détenant uniquement un téléphone cellulaire ou encore les mises à jour récentes ne figurent pas sur la base liste. La base aréolaire et la base à CA compensent cette lacune.

**Tableau 2.1.3 : Couverture téléphonique (%) des régions sociosanitaires présentée selon la province**

Province	Moyenne	Minimum	Maximum
Terre-Neuve	98,1	97,2	98,6
Île-du-Prince-Édouard	98,0	97,1	98,9
Nouvelle-Écosse	97,7	96,2	99,3
Nouveau-Brunswick	98,5	98,0	99,3
Québec	98,3	95,9	99,3
Ontario	98,9	96,7	100,0
Manitoba	98,3	96,1	99,5
Saskatchewan	99,0	97,8	99,7
Alberta	99,1	97,6	100,0
Colombie-Britannique	98,4	95,0	100,0
<b>Canada</b>	<b>98,6</b>	<b>95,0</b>	<b>100,0</b>

**Tableau 2.1.4 : Couverture de la base de numéros de téléphone (%) des**

### régions sociosanitaires présentée selon la province

Province	Moyenne	Minimum	Maximum
Terre-Neuve	81	75	88
Île-du-Prince-Édouard	87	86	88
Nouvelle-Écosse	83	83	83
Nouveau-Brunswick	-	-	-
Québec	80	74	86
Ontario	77	58	86
Manitoba	76	62	86
Saskatchewan	88	83	93
Alberta	75	54	84
Colombie-Britannique	74	60	87
<b>Canada</b>	<b>80</b>	<b>54</b>	<b>93</b>

Le *taux de vacance* représente le pourcentage de logements, parmi les logements sélectionnés, qui sont vacants au moment de la collecte sur le terrain. Les logements vacants font partie des unités hors du champ de l'enquête. Ce type de logement, lorsque identifié correctement, n'introduit aucun biais dans les estimations de l'ESCC. Par contre, la variance de l'estimation s'en trouve plus élevée puisque l'échantillon compte un nombre moins élevé de ménages. Lors de la planification de l'ESCC, ce taux a été pris en considération afin d'obtenir les tailles espérées et par le fait même réduire la variance.

Une attention particulière doit donc être apportée à la détermination des logements vacants puisque ceux-ci influencent directement deux autres indicateurs de qualité : le taux de glissement et le taux de réponse qui sera discuté à la section 2.2. En effet, si un logement est codé vacant plutôt que d'être identifié comme étant temporairement absent par exemple, le taux de non-réponse produit par l'ESCC est quelque peu sous-estimé. Par ailleurs, le taux de glissement s'en trouve surestimé puisque ce logement mal codé n'aurait pas dû être considéré comme une unité hors du champ de l'enquête. Le tableau 2.1.5 présente le taux de vacance moyen selon les deux types de strates utilisées dans l'enquête (urbaine et rurale). Les taux nationaux pour la partie urbaine et rurale s'établissent à 8,8 % et 27,8 % respectivement. Ces taux sont très comparables à ceux observés à l'EPA et à ceux utilisés lors de la planification de l'échantillon. De plus, l'échantillon de l'ESCC n'étant pas choisi en une seule occasion, il est possible de réajuster la taille de l'échantillon selon les taux observés tout au long de la collecte.

**Tableau 2.1.5 : Taux de vacance moyen selon les strates (urbaine/rurale) par province – Base aréolaire**

Province	Urbain	Rural	Total
Terre-Neuve	11,4	32,0	20,4
Île-du-Prince-Édouard	10,8	30,4	21,2
Nouvelle-Écosse	10,9	25,4	17,8
Nouveau-Brunswick	13,6	20,2	16,6
Québec	8,9	28,6	13,9
Ontario	7,6	30,7	11,8
Manitoba	7,9	35,5	19,8
Saskatchewan	7,2	28,4	17,1
Alberta	7,8	18,7	10,5
Colombie-Britannique	8,5	21,1	10,4
Yukon	18,4	-	18,4
T.N.-O.	22,9	-	22,9
Nunavut	17,1	-	17,1
<b>Canada</b>	<b>8,8</b>	<b>27,8</b>	<b>13,9</b>



## 2.2 Suivi des opérations de collecte

Les tableaux 2.2.1 et 2.2.2 présentent les tailles d'échantillon espérées et celles observées lors de la collecte selon le type de base de sondage pour les provinces/territoires et par groupe d'âge à l'échelle du Canada.

Puisqu'il a été possible de réajuster la taille de l'échantillon selon les taux observés tout au long de la collecte, les tailles observées sont assez proches de celles espérées. Pour la base aréolaire, en général les tailles observées sont légèrement plus petites que celles espérées alors que c'est l'inverse pour la base téléphonique où elles sont légèrement plus grandes. On remarque toutefois deux exceptions : l'Île-du-Prince-Édouard et l'Ontario.

Dans le cas de l'Île-du-Prince-Édouard, un achat de 1 300 unités d'échantillonnage supplémentaires a été fait au cours de la collecte des données dans le but d'obtenir un échantillon de taille suffisante pour produire des estimations fiables pour cinq régions infraprovinciales au lieu de deux comme il en était question dans le plan d'échantillonnage original. L'échantillon d'unités supplémentaires a été sélectionné entièrement à partir de la base de numéros de téléphone.

En Ontario, pour des raisons de charge de travail accumulée au cours des premiers mois de collecte et des problèmes de capacités techniques, le nombre de ménages sélectionnés à partir de la base aréolaire a dû être réduit considérablement après le premier trimestre de collecte. Cette diminution a permis aux intervieweurs de reprendre le dessus et aux gestionnaires d'engager et de former du nouveau personnel. Afin de minimiser l'incidence de cette diminution d'interviews personnelles sur l'échantillon total, un plus grand nombre d'interviews téléphoniques ont dû être faites.

**Tableau 2.2.1 : Taille de l'échantillon espérée versus observée selon la base de sondage (en nombre de personnes)**

Province	Base aréolaire		Base téléphonique		Échantillon total	
	espérée	observée	espérée	observée	espérée	observée
Terre-Neuve	3 830	3 596	180	274	4 010	3 870
Île-du-Prince-Édouard	1 810	1 722	190	1 929	2 000	3 651
Nouvelle-Écosse	4 540	4 686	500	633	5 040	5 319
Nouveau-Brunswick	5 030	4 863	120	133	5 150	4 996
Québec	22 100	20 482	2 180	2 185	24 280	22 667
Ontario	35 855	30 853	6 405	8 425	42 260	39 278
Manitoba	7 765	7 697	235	773	8 000	8 470
Saskatchewan	7 160	6 949	560	1 060	7 720	8 009
Alberta	11 605	10 935	2 595	3 521	14 200	14 456
Colombie-Britannique	14 835	13 899	3 255	4 403	18 090	18 302
Yukon	757	635	94	174	850	809
T.N.-O.	810	919	90	82	900	1 001
Nunavut	720	670	80	37	800	707
<b>Canada</b>	<b>116 817</b>	<b>107 906</b>	<b>16 484</b>	<b>23 629</b>	<b>133 300</b>	<b>131 535</b>

À l'échelle nationale, la distribution des tailles d'échantillon observées par groupe d'âge (tableau 2.2.2) s'apparente à celle des tailles espérées. On peut séparer l'échantillon total en deux groupes : dans le premier on trouve les 12-19 ans, les 20-29 ans et les 65 ans et plus qui représentent respectivement 13,4 %, 12,5 % et 18,5 % de l'échantillon total alors que dans le deuxième on trouve les 30-44 ans et les 45-64 ans qui représentent un peu plus de 55 % de l'échantillon total. Cette observation concorde avec les données démographiques du recensement. Il faut rappeler ici que la sélection des répondants individuels a été conçue de sorte que les jeunes (de 12 à 19 ans) et les personnes âgées (65 ans et plus) soient bien représentés dans l'échantillon.

**Tableau 2.2.2 : Distribution par groupe d'âge : Taille de l'échantillon  
espérée versus observée selon la base de sondage (%)**

Âge	Base aréolaire		Base téléphonique		Échantillon total	
	espérée	observée	espérée	observée	espérée	observée
12-19	15,9	14,4	8,5	9,1	12,5	13,4
20-29	12,7	12,6	14,3	12,3	13,4	12,5
30-44	28,2	27,2	29,1	27,4	28,6	27,3
45-64	26,0	27,9	28,0	30,9	26,9	28,4
65 +	17,3	18,1	20,2	20,3	18,6	18,5

Tout au long de la collecte, la non-réponse a été analysée de très près. En effet, l'ampleur de la non-réponse varie de non-réponse partielle (le fait de ne pas répondre à une ou plusieurs questions) à une non-réponse totale. À l'ESCC, il n'y a presque pas de non-réponse partielle car une fois le questionnaire débuté les répondants avaient tendance à le terminer. Il y a eu non-réponse totale lorsque le ménage n'a pas répondu, lorsque la personne sélectionnée pour participer à l'enquête a refusé de le faire, ou que l'intervieweur a été incapable d'entrer en contact avec elle. On a traité les cas de non-réponse totale en corrigeant tout d'abord le poids à l'échelle du ménage et ensuite les poids des personnes qui ont répondu à l'enquête afin de compenser pour celles qui n'ont pas répondu (Brisebois et Thivierge, 2001).

Pour l'échantillon total, des taux de réponse de 91,4 % et 91,9 % ont été obtenus aux échelles du ménage et de la personne (tableau 2.2.3). Pour l'ESCC, on a observé un taux de réponse combiné de près de 85 %. Par ailleurs, des taux de réponse combinés d'environ 85 % et 83 % ont été obtenus pour les bases aréolaire et téléphonique à l'échelle nationale. Ces taux sont légèrement inférieurs à ceux utilisés lors de la planification de l'échantillon. De plus, les taux provinciaux et territoriaux varient entre 66,3 % (Nunavut) et 90,0 % (Manitoba) pour la base aréolaire alors qu'ils varient entre 34,6 % (Nunavut) et 95,6 % (Yukon) pour la base téléphonique. Pour les deux bases de sondage, c'est au Nunavut qu'on obtient le plus faible taux de réponse. Des raisons opérationnelles, conjuguées avec la nouveauté d'effectuer une enquête au Nunavut, expliquent ce faible taux.

**Tableau 2.2.3 : Taux de réponse à l'échelle de la personne et du ménage selon le type de base**

Province	Base aréolaire			Base téléphonique			Échantillon total		
	Ménage	Personne	Combiné	Ménage	Personne	Combiné	Ménage	Personne	Combiné
Terre-Neuve	94,6	90,8	86,6	96,1	92,9	89,3	94,7	91,0	86,8
Île-du-Prince-Édouard	93,6	92,7	87,7	88,8	93,0	82,6	90,8	92,9	84,7
N.-Écosse	93,9	93,6	88,8	94,5	94,5	89,3	94,0	93,7	88,8
N.-Brunswick	94,1	93,0	88,4	97,9	94,3	92,4	94,2	93,1	88,5
Québec	92,0	92,5	85,7	90,8	93,3	84,8	91,9	92,6	85,6
Ontario	90,9	89,9	82,8	86,4	92,1	79,5	89,8	90,4	82,0
Manitoba	94,3	94,8	90,0	90,6	93,8	85,0	93,9	94,7	89,5
Saskatchewan	92,7	92,9	87,0	92,9	91,9	85,3	92,7	92,8	86,8
Alberta	93,1	90,5	85,2	92,7	91,6	84,9	93,0	90,7	85,1
C.-Britannique	89,8	92,7	83,9	92,9	93,4	86,7	90,6	92,8	84,7
Yukon	85,2	91,9	79,3	97,3	98,3	95,6	87,7	93,2	82,7
T.N.-O.	96,2	92,4	89,6	93,8	91,1	85,4	96,0	92,3	89,2
Nunavut	74,1	86,6	66,3	37,4	92,5	34,6	69,6	86,9	62,5
<b>Canada</b>	<b>91,8</b>	<b>91,7</b>	<b>85,1</b>	<b>89,8</b>	<b>92,6</b>	<b>83,1</b>	<b>91,4</b>	<b>91,9</b>	<b>84,7</b>

À titre indicatif, le tableau 2.2.4 présente les taux de réponse combinés minimaux et maximaux par région sociosanitaire selon la province et la base de sondage. Les taux de réponse combinés par RSS

oscillent entre 66,3 % et 96,7 % pour l'échantillon de la base aréolaire, entre 34,6 % et 100,0 % pour celui de la base téléphonique et entre 62,5 % et 94,9 % pour l'échantillon total.

**Tableau 2.2.4 : Taux de réponse combiné par région sociosanitaire présentés selon la province et la base de sondage**

Province	Base aréolaire		Base téléphonique		Échantillon total	
	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
Terre-Neuve	79,2	96,7	88,8	89,9	84,7	92,5
Île-du-Prince-Édouard	87,5	87,7	82,4	83,5	83,9	86,1
Nouvelle-Écosse	86,2	92,3	87,3	100,0	86,2	90,6
Nouveau-Brunswick	86,4	93,6	90,1	100,0	87,8	90,0
Québec	70,7	91,2	75,2	89,5	73,4	91,2
Ontario	70,2	93,5	55,6	89,7	70,2	89,3
Manitoba	86,7	95,1	79,4	89,3	86,7	94,9
Saskatchewan	81,8	92,8	88,4	96,6	83,4	91,5
Alberta	76,0	90,6	79,5	100,0	80,9	90,6
Colombie-Britannique	76,3	95,6	81,5	91,8	76,8	90,5
Yukon	79,3	79,3	95,6	95,6	82,7	82,7
T.N.-O.	89,6	89,6	85,4	85,4	89,2	89,2
Nunavut	66,3	66,3	34,6	34,6	62,5	62,5
<b>Canada</b>	<b>66,3</b>	<b>96,7</b>	<b>34,6</b>	<b>100,0</b>	<b>62,5</b>	<b>94,9</b>

Lors de la sélection des répondants, deux personnes de 12 ans et plus ont été sélectionnées au hasard uniquement dans 18 % des ménages échantillonnés à partir de la base. Cette approche a été adoptée dans l'optique de minimiser le fardeau de réponse au téléphone. Les taux de réponse à l'échelle de la personne en fonction du nombre de personnes sélectionnées dans le ménage selon la base de sondage sont présentés dans le tableau 2.2.5.

La charge de réponse étant plus élevée lorsque deux personnes sont sélectionnées dans un même ménage, le taux de réponse à l'échelle de la personne est plus faible que lorsqu'une seule personne est sélectionnée et ce, pour toutes les provinces/territoires. Pour le Canada, près de 10 points de pourcentage séparent ces deux taux. En effet, il est d'environ 86 % lorsque deux répondants sont sélectionnés alors qu'il est de 94 % lorsqu'il n'y a qu'un seul répondant. Par ailleurs, lorsqu'une seule personne est sélectionnée dans le ménage, les taux de réponse pour les bases aréolaire et téléphonique sont tous supérieurs à 90 %. Ceux de la base aréolaire sont légèrement supérieurs à ceux de la base téléphonique à l'exception de celui du Yukon.

**Tableau 2.2.5 : Taux de réponse à l'échelle de la personne en fonction du nombre de personnes sélectionnées dans le ménage selon la base de sondage par province**

Province	2 personnes sélectionnées	1 personne sélectionnée	
	Base aréolaire	Base aréolaire	Base téléphonique
Terre-Neuve	87,5	92,6	92,9
Île-du-Prince-Édouard	87,5	94,9	93,0
Nouvelle-Écosse	88,6	95,7	94,5
Nouveau-Brunswick	87,5	95,3	94,3
Québec	87,9	94,1	93,3
Ontario	83,5	92,7	92,1
Manitoba	90,9	96,3	93,8
Saskatchewan	86,9	95,4	91,9
Alberta	84,8	93,0	91,6
Colombie-Britannique	88,1	94,5	93,4

Yukon	85,2	94,8	98,3
T.N.-O.	88,9	94,3	91,1
Nunavut	77,9	95,5	92,5
<b>Canada</b>	<b>86,3</b>	<b>94,0</b>	<b>92,6</b>

Différents moyens ont été mis en place pour tenter d'éviter la non-réponse tels l'envoi de lettres expliquant l'enquête, la prise de rendez-vous, les visites nombreuses, la traduction du questionnaire en plusieurs langues, etc. Malgré cela, une partie des ménages sélectionnés pour faire partie de l'ESCC s'est soldée en non-réponse et ce, pour différentes raisons. Les distributions de la non-réponse à l'échelle du ménage selon le type de non-réponse par province/territoire pour les bases de sondage aréolaire et téléphonique sont présentées dans les tableaux 2.2.6 et 2.2.7. Pour la base aréolaire et la base téléphonique, on observe le même scénario. À l'échelle canadienne, plus de 60 % des cas de non-réponse sont attribuables à un refus de participer à l'enquête. Les catégories « personne à la maison » et « autres » représentent un peu plus de 30 % des cas de non-réponse alors que les raisons d'absence pour la durée de l'enquête et de barrière linguistique représentent ensemble environ 4% de la non-réponse totale.

**Tableau 2.2.6 : Distribution de la non-réponse à l'échelle du ménage selon le type de non-réponse par province - Base aréolaire**

Province	Nombre de non-réponse	Personne à la maison	Absent pour la durée de l'enquête	Barrière linguistique	Refus	Autres
Terre-Neuve	186	10,8	4,3	0,0	66,7	18,3
Île-du-Prince-Édouard	108	10,2	2,8	1,9	73,1	12,0
Nouvelle-Écosse	275	22,9	2,5	1,5	60,4	12,7
Nouveau-Brunswick	282	9,2	1,4	0,0	75,2	14,2
Québec	1 683	19,0	2,4	2,7	63,3	12,7
Ontario	2 907	25,2	1,6	1,9	61,5	9,8
Manitoba	419	14,1	1,7	2,4	71,8	10,0
Saskatchewan	506	11,1	2,6	0,4	77,5	8,5
Alberta	760	17,0	3,6	1,8	70,1	7,5
Colombie-Britannique	1 457	18,2	2,5	3,2	63,7	12,5
Yukon	102	18,6	7,8	0,0	53,9	19,6
T.N.-O.	32	31,3	0,0	3,1	46,9	18,8
Nunavut	202	26,2	10,9	4,5	8,9	49,5
<b>Canada</b>	<b>8 919</b>	<b>19,8</b>	<b>2,5</b>	<b>2,1</b>	<b>63,6</b>	<b>12,0</b>

**Tableau 2.2.7 : Distribution de la non-réponse à l'échelle du ménage selon le type de non-réponse par province - Base téléphonique**

Province	Nombre de non-réponse	Personne à la maison	Absent pour la durée de l'enquête	Barrière linguistique	Refus	Autres
Terre-Neuve	12	0,0	0,0	0,0	91,7	8,3
Île-du-Prince-Édouard	261	11,5	2,3	0,4	61,7	24,1
Nouvelle-Écosse	202	9,9	3,0	0,0	62,9	24,3
Nouveau-Brunswick	3	0,0	0,0	0,0	100,0	0,0
Québec	236	16,9	0,8	0,4	57,6	24,2
Ontario	1 442	19,0	1,4	2,8	66,8	10,0
Manitoba	85	2,4	1,2	0,0	85,9	10,6
Saskatchewan	88	6,8	1,1	0,0	85,2	6,8
Alberta	304	9,9	1,0	0,3	72,4	16,4
Colombie-Britannique	18	11,1	0,0	0,0	83,3	5,6
Yukon	5	0,0	0,0	0,0	80,0	20,0
T.N.-O.	6	0,0	16,7	0,0	83,3	0,0
Nunavut	67	82,1	1,5	1,5	4,5	10,4
<b>Canada</b>	<b>2 909</b>	<b>16,1</b>	<b>1,3</b>	<b>2,3</b>	<b>66,3</b>	<b>14,1</b>

Les tableaux 2.2.8 et 2.2.9 permettent de dresser un profil des répondants et des non-répondants par groupe d'âge selon la base de sondage et par groupe d'âge-sexe à l'échelle du Canada. On remarque tout d'abord une différence entre la distribution des non-répondants et des répondants, principalement pour les groupes d'âge 12-19 ans et les 65 ans et plus (tableau 3.2.8). Pour les répondants, une différence marquée est observée entre les bases aréolaire et téléphonique pour les 12-19 ans (14,3 % versus 9,1 respectivement). Pour les non-répondants, la différence la plus notable est présente chez les 12-19 ans (30,8 % pour la base aréolaire versus 9,1 pour la base téléphonique) suivie des 45-64 ans (22,9 % versus 31,1 respectivement). D'un autre côté, la distribution des répondants et des non-répondants par groupe d'âge-sexe (tableau 2.2.9) affiche une différence plus marquée chez les 65 ans et plus, dans les deux cas les hommes affichent des taux inférieurs de près de 5 points de pourcentage à ceux des femmes.

**Tableau 2.2.8 : Distribution des répondants et des non-répondants par groupe d'âge selon la base de sondage - Canada**

Âge	Répondants			Non-répondants		
	aréolaire	téléphonique	total	aréolaire	téléphonique	total
12-19	14,3	9,1	13,4	30,8	9,1	27,3
20-29	12,6	12,3	12,5	11,1	12,4	11,4
30-44	27,2	27,4	27,3	24,2	28,1	24,8
45-64	27,8	30,9	28,4	22,9	32,1	24,1
65 +	18,1	20,3	18,5	10,9	18,3	12,1

**Tableau 2.2.9 : Distribution des répondants et des non-répondants par groupe d'âge-sexe - Canada**

Âge	Répondants			Non-répondants		
	Homme	Femme	total	Homme	Femme	total
12-19	14,6	12,4	13,4	26,3	28,6	27,3
20-29	12,4	12,6	12,5	11,7	10,9	11,4
30-44	27,9	26,7	27,3	26,0	23,4	24,8
45-64	29,3	27,6	28,4	25,9	22,5	24,1
65 +	15,9	20,7	18,5	10,1	14,8	12,1

## 2.3 Certification des données collectées

Tout au long de la collecte, les données recueillies ont été mises à l'épreuve lors de l'étape de certification des données. Dans un premier temps, la durée des interviews a fait l'objet d'une analyse approfondie. Une durée moyenne d'interview a été déterminée et un temps improbable a été établi c'est-à-dire une durée en deçà de laquelle il était impossible de réaliser une interview. Dans un deuxième temps, plusieurs statistiques ont été obtenues pour chaque intervieweur. On a observé que quelques intervieweurs réalisaient de façon systématique des interviews d'une longueur improbable. En présence d'un comportement douteux, une deuxième interview a été réalisée par un intervieweur principal. Cette pratique a permis de congédier un faible nombre d'intervieweurs ne répondant pas aux attentes de l'Agence. Par la suite, une analyse a permis d'évaluer le biais introduit dans les estimations afin d'évaluer l'incidence des interviews jugées douteuses. Ce processus de certification a mené à la décision d'éliminer 3 847 enregistrements de la base de données finale.

## 2.4 Incidence du mode de collecte (personnelle, téléphone) et de la forme de déclaration (auto-déclarée ou par procuration) sur les réponses

Selon plusieurs études antérieures effectuées au Canada et aux États-Unis, des différences significatives entre les estimations de certaines caractéristiques reliées à la santé peuvent être observées selon le mode de collecte (interview en personne ou téléphonique) et la forme de déclaration (auto-déclaration ou par procuration). Par exemple, plusieurs études ont rapporté que les interviews téléphoniques produisent des estimations de la consommation d'alcool ou de drogues et d'inactivité physique plus faibles que les interviews en personne (effet du mode de collecte). Également, d'autres études rapportent que les prévalences des problèmes de santé de longue durée tels que les maladies chroniques, la limitation des activités, l'hospitalisation et les problèmes émotionnels calculées à partir des données recueillies par procuration seraient inférieures à celles obtenues à partir des données auto-déclarées (effet de la forme de déclaration).

Pour évaluer l'incidence du mode de collecte, on a réalisé une analyse comparative unidimensionnelle des estimations ponctuelles avec un test statistique de signification unilatérale, puis une analyse multidimensionnelle où la régression logistique a été utilisée. Afin d'éliminer l'élément « base de sondage » de l'analyse, seule la base aréolaire a été retenue pour cette étude (dont 58 % des interviews ont été réalisées en personne et 42 % par téléphone). Plus de détails sur cette étude sont disponibles dans Pierre et Béland (2002).

Le tableau 2.4.1 présente certains résultats de la première analyse; des différences significatives ont été observées entre les deux modes de collecte pour plusieurs indicateurs de la santé. Les interviews téléphoniques ont produit par exemple des estimations statistiquement plus basses que les interviews en personne pour les caractéristiques suivantes : limitation des activités ou handicap, la déclaration d'au moins un problème de santé chronique, hospitalisation et inactivité physique. Par contre, les estimations des consultations d'un intervenant en médecine douce et de la consommation d'alcool sont statistiquement plus élevées pour les interviews téléphoniques. De plus, les résultats de cette étude démontrent que l'effet de la déclaration par procuration est indéniable. La propension à déclarer certains problèmes de santé est notamment plus faible pour les réponses obtenues par procuration que pour celles auto-déclarées. Afin d'atténuer le plus possible ce phénomène, les procédures de collecte pour les cycles futurs de l'ESCC ont été clarifiées et renforcées.

Afin de mieux comprendre ces résultats, une étude plus approfondie sera réalisée au deuxième cycle de vie de l'enquête dans un environnement plus contrôlé. Le plan d'échantillonnage actuel de l'ESCC n'a pas été défini de façon à effectuer une analyse « contrôlée » du mode de collecte; il n'y a donc pas eu de randomisation pour assurer une bonne représentativité de chaque sous-échantillon (téléphonique versus en personne). Il est donc important d'interpréter ces résultats avec prudence.

**Tableau 2.4.1 : Estimations de la prévalence de certains problèmes de santé selon le mode de collecte  
Base aréolaire - Canada**

Problème de santé	Mode de collecte	
	Personnelle	Téléphonique
Consultation d'un médecin	81,9	82,2
Limitation des activités	32,8*	26,7
Maux de dos	17,5	17,2
Limitations des activités /2	15,8	16,3
C.P. de la santé	8,7	8,8
Taux de fumeur	25,9	26,3
Problème de santé	65,8*	61,4
Risque de dépression	10,2	10,6
C.I. en médecine douce	10,3**	11,7
Hospitalisation	9,2*	6,9
Inactivité physique	57,0*	51,8
Alcool : cons. modérée	21,6**	24,8
Alcool : cons. élevée	19,3**	21,3

**Nota :** Les estimations de la prévalence ont été calculées en excluant les cas pour lesquels des valeurs manquaient. La proportion d'enregistrements avec valeurs manquantes est inférieure à un quart de point de pourcentage pour chaque indicateur de santé. La comparaison des estimations se fonde sur un test de signification unilatéral.

\* : valeur significativement plus élevée que l'estimation calculée pour les interviews téléphoniques,  $p < 0,05$

\*\* : valeur significativement plus faible que l'estimation calculée pour les interviews téléphoniques,  $p < 0,05$

## 2.5 Indicateurs d'imputation

Étant donné le caractère privé ou délicat de l'ESCC, de nombreuses questions ou modules du questionnaire ne se prêtaient qu'à des réponses personnelles et n'étaient pas posées lorsque l'interview était effectuée par procuration. Au cours de la collecte des données, une plus grande proportion d'interviews que prévue a été faite par procuration bien que des instructions claires aient été données aux intervieweurs. Après les six premiers mois de collecte, le taux d'interviews par procuration national atteignait la barre des 7,5 %. Bien que les instructions aux intervieweurs aient été renforcées pour rectifier la situation, à la fin de la collecte des données, 6,3 % des interviews avaient été complétées par procuration (voir tableau 2.5.1). Le taux variait entre 2 % et 23 % à l'échelle des régions sociosanitaires. En présence d'interviews par procuration, environ le tiers du questionnaire n'était pas complété.

**Tableau 2.5.1 : Taux d'interviews par procuration par province**

Province	Taux d'interviews par procuration
Terre-Neuve	4,9
Île-du-Prince-Édouard	4,2
Nouvelle-Écosse	3,9
Nouveau-Brunswick	9,1
Québec	5,2
Ontario	5,4
Manitoba	8,3
Saskatchewan	5,9
Alberta	5,9
Colombie-Britannique	9,5
Yukon	8,4
T.N.-O.	12,7
Nunavut	6,9
<b>Canada</b>	<b>6,3</b>

Pour pallier les données manquantes suite au nombre élevé d'interviews réalisées par procuration, la méthode d'imputation « du plus proche voisin » a été adoptée (St-Pierre et Béland, 2002). L'imputation n'a été utilisée que pour compléter les renseignements recueillis par procuration. Elle n'a pas été utilisée en présence de non-réponse totale ou partielle lors d'interviews réalisées auprès de la personne sélectionnée. En vertu de cette méthode, on a recherché, dans des catégories d'imputation prédéfinies, un répondant « donneur » dont les caractéristiques étaient semblables (le plus proche voisin) à celles de la personne pour laquelle les renseignements avaient été recueillis par procuration ; puis, on a imputé à cette dernière les valeurs recueillies pour le « donneur ». On a trouvé le plus proche voisin à l'aide d'une fonction de distance précise qui utilisait des renseignements pertinents disponibles à la fois sur le répondant par procuration et sur le répondant « donneur ». Là où on ne pouvait assurer la qualité des données par imputation (selon des critères pré-établis), on a attribué le code manquant aux réponses.

À titre indicatif, le tableau 2.5.2 présente quelques taux d'imputation par province/territoire pour quatre modules du questionnaire : tension artérielle, utilisation des services de santé mentale, consommation de fruits et légumes et dépression. À l'échelle nationale, pour ces modules, les taux d'imputation se situent aux alentours de 6%. Pour ces quatre modules, les taux les plus bas ont été observés en Nouvelle-Écosse tandis que les plus élevés ont été enregistrés aux T.N.-O ; ce qui laisse présager qu'il s'agissait bel et bien d'un enjeu de collecte.



**Tableau 2.5.2 : Taux d'imputation pour quatre modules par province**

Province	Tension artérielle	Utilisation des services de santé mentale	Consommation de fruits et de légumes	Dépression
Terre-Neuve	4,9	4,7	4,9	4,7
Île-du-Prince-Édouard	4,2	4,0	4,2	4,0
Nouvelle-Écosse	3,9	3,8	3,9	3,8
Nouveau-Brunswick	8,9	8,7	9,0	8,7
Québec	5,2	5,1	5,2	5,1
Ontario	5,3	5,2	5,3	5,2
Manitoba	8,2	8,2	8,3	8,2
Saskatchewan	5,4	5,1	5,6	5,4
Alberta	5,8	5,8	5,8	5,8
Colombie-Britannique	9,4	9,3	9,4	9,3
Yukon	7,7	7,3	8,2	7,3
T.N.-O.	10,0	9,7	11,7	9,7
Nunavut	5,5	6,1	6,5	6,1
<b>Canada</b>	<b>6,2</b>	<b>6,1</b>	<b>6,2</b>	<b>6,1</b>

## 2.6 Erreurs dues à l'échantillonnage

Étant donné que les estimations d'une enquête comportent inévitablement une erreur due à l'échantillonnage, l'ESCC fournit à ses utilisateurs une mesure de variabilité pour une multitude d'estimation de type nominal, en l'occurrence des tableaux de coefficients de variation (CV). Les CV ne sont pas présentés dans cet article mais ont été analysés attentivement à l'ESCC.

Une autre mesure de qualité analysée à l'ESCC est *l'effet du plan de sondage*. Cet indicateur de qualité donne une idée de l'efficacité du plan de sondage combiné à la méthode d'estimation. Il est possible de dériver plusieurs effets de plan de sondage pour une vaste gamme de caractéristiques mesurées à l'ESCC pour un grand nombre de domaines d'intérêt (groupes âge-sexe par exemple). Le tableau 3.6.1 présente les effets de plan de sondage minimaux et maximaux pour l'ensemble des RSS à l'intérieur d'une province et par province ou territoire. Les effets de plan présentés oscillent entre 1,12 et 2,21 pour les régions sociosanitaires et entre 1,12 et 2,44 pour les provinces/territoires. Il se positionne à 2,34 pour le Canada.

**Tableau 2.6.1 : Effets de plan de sondage minimaux, maximaux et global pour les RSS présentés par province**

Province	Global	Minimum	Maximum
Terre-Neuve	1,63	1,37	1,93
Île-du-Prince-Édouard	2,02	1,99	2,05
Nouvelle-Écosse	1,63	1,39	1,63
Nouveau-Brunswick	1,51	1,35	1,52
Québec	2,08	1,44	2,21
Ontario	2,31	1,39	2,11
Manitoba	2,44	1,54	2,01
Saskatchewan	1,82	1,35	1,75
Alberta	2,39	1,44	2,09
Colombie-Britannique	1,77	1,39	1,69
Territoires	1,12	1,12	1,12
<b>Canada</b>	<b>2,34</b>	<b>1,12</b>	<b>2,21</b>

## **Mot de la fin**

Tout au long du déroulement de la collecte, un système de gestion de la qualité a permis de contrôler la qualité des données de l'ESCC et de réagir le cas échéant (Béland, Dufour et Hamel, 2001). Puisque l'ESCC était à ses premières armes, il est évident que ce système de gestion puisse bénéficier d'améliorations pour les réalisations futures de l'ESCC. Plusieurs autres analyses devront également être réalisées afin de mieux comprendre le premier cycle de cette enquête.

De plus, à la lumière de cette première réalisation, il est également évident que des modifications seront apportées aux prochains cycles de l'enquête. Par exemple, un échantillon contrôlé permettrait d'analyser plus efficacement l'incidence du mode de collecte (en personne ou téléphonique) sur les estimations. On pense également à mettre sur pied un programme de validation des intervieweurs administré d'une façon plus centrale. On envisage aussi de sélectionner uniquement une seule personne par ménage pour les autres cycles. On a également constaté que la liste de numéros de téléphone utilisée comme base d'échantillonnage bénéficierait d'une mise à jour plus fréquente.

Les prochaines étapes pour l'équipe de projet consisteront à poursuivre les différentes études entreprises et à s'interroger sur les améliorations possibles pour le futur.

## **REMERCIEMENTS**

Les auteurs tiennent à souligner l'excellent travail des méthodologistes suivants durant le déroulement de l'étude d'évaluation de la qualité des données de l'ESCC : François Brisebois, Marco Grenier, Suzanne Lessard, Marianna Morano, Fritz Pierre, Martin St-Pierre et Sylvain Thivierge. Ils tiennent également à remercier France Bilocq et Johanne Tremblay pour leurs précieux commentaires lors de la rédaction de cet article.

## BIBLIOGRAPHIE

Béland, Y., Bailie, L., Catlin, G. et Singh, M.P. (2000). CCHS and NPHS — An Improved Health Survey Program at Statistics Canada, *2000 Proceedings of the American Statistical Association Meeting, Survey Research Methods Section*.

Béland, Y., Dufour, J. et Hamel, M. (2001). Faire échec à la non-réponse dans le cadre de l'Enquête sur la santé dans les collectivités canadiennes, Actes du Symposium 2001, Statistique Canada.

Brisebois, F. et Thivierge, S. (2001). The Weighting Strategy for the Canadian Community Health Survey, Statistique Canada, *2001 Proceedings of the American Statistical Association Meeting, Survey Research Methods Section*.

Institut canadien de l'information sur la santé (1999). *Carnet de route de l'information sur la santé : Entreprendre le voyage*. (1-895581-32-X).

Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B., et Lindsey, J. (1998). *Méthodologie de l'Enquête sur la population active du Canada*, Statistique Canada no. 71-526-XPB au catalogue.

Norris, D.A. et Paton, D.G. (1991). L'Enquête sociale générale canadienne : bilan des cinq premières années. *Techniques d'enquête*, 17, p : 245-260.

Pierre, F. et Béland, Y. (2002). Étude sur quelques erreurs de réponse dans le cadre de l'ESCC – Recueil des présentations de la section des méthodes d'enquête, Société statistique du Canada; p : 69-75.

Statistique Canada (1998). *Lignes directrices concernant la qualité*. Troisième Édition, octobre 1998, Statistique Canada no. 12-539-X1F au catalogue.

St-Pierre, M. et Béland, Y. (2002). *Imputation of proxy respondents in the Canadian Community Health Survey* – Recueil des présentations de la section des méthodes d'enquête, Société statistique du Canada; p : 45-50.